SIGNIFICANCE

Physical methods are poised to transform drug discovery and chemical biology by enabling true molecular design. While modeling work is used extensively in drug discovery, its main role at present is to aid with idea generation or to filter large libraries of compounds for screening. Instead, we imagine using computational techniques extensively to guide the design process. Consider a medicinal chemist in the not-too-distant future who has just finished synthesizing several new derivatives of an existing inhibitor as potential drug leads targeting a particular biomolecule, and has obtained binding affinity or potency data against the desired biomolecular target. Before leaving work, he or she generates ideas for perhaps 100 new compounds which could be synthesized next, then sets a computer to work overnight prioritizing them. By morning, the compounds have all been prioritized based on reliable predictions of their affinity for the desired target, selectivity against alternative targets which should be avoided, solubility, and membrane permeability. The chemist then looks through the predicted properties for the top few compounds and selects the next ones for synthesis. If synthesizing and testing each compound takes several days, this workflow compresses roughly a year's work into a few days.

While this workflow is not yet a reality, huge strides have been made in this direction, with calculated binding affinity predictions now showing real promise [1–8], rapid progress toward solubility predictions [9–11], and selectivity and drug resistance also apparently tractable [12], with some headway apparent on membrane permeability [13, 14]. A considerable amount of science and engineering still remains to make this vision a reality, but **given recent progress, the question now seems more one of *when* rather than *whether*.**

The widespread availability of inexpensive graphics processing units (GPUs), which provide a 100-fold increase in price/performance over CPUs, coupled with advances in automation [15] and sampling protocols have helped simulation-based techniques reach the point where they now begin to be genuinely useful in guiding drug discovery for a limited *domain of applicability* [4–8, 16, 17]. Specifically, in some situations, free energy calculations appear to be capable of achieving RMS errors of 1-2 kcal/mol with current force fields, even in prospective applications, sufficient to drastically reduce the number of molecules that must be synthesized and assayed [18]. As a consequence, pharmaceutical companies are beginning to use these methods in active discovery projects.

Despite this progress, these methods currently have severe limitations, and a great deal of science and engineering is still needed before these techniques can achieve their full potential in transforming molecular design. For example, even "small" protein conformational conformational changes not gracefully handled by current methodologies can yield errors up to 5 kcal/mol in calculated binding free energies [19], force field limitations still pose major challenges [20], and the inability to treat important chemical effects like protonation and tautomer equilibria drastically limits the domain of applicability. For many pharmaceutically relevant systems, the most important sources of error are not yet clear.

Progress on addressing these challenges has been frustratingly slow, hindered by a lack of high-quality data and community focus. **Neither retrospective tests nor prospective application in active drug discovery projects provides the necessary impetus and data to rapidly overcome the remaining major barriers to widespread utility.** Large-scale retrospective tests certainly assess our *retrospective* performance, but they do not provide accurate guidance on utility for prospective design, nor do they effectively pave the way forward. One issue with retrospective tests is that they can easily result in over-fit models, or researchers applying a variety of different protocols until they get good results, perhaps even by a statistical fluke [ref] . In retrospective tests, performance may also not be indicative of expected performance in applications because even well-meaning researchers can take advantage of prior knowledge. For example, if the binding mode of a ligand is already known crystallographically, a researcher may use that binding mode in retrospective tests, whereas prospective or design work would require first selecting among candidate binding modes, introducing substantial uncertainty unaccounted for in the retrospective statistics [1, 21, 22]. This also means that in retrospective tests, researchers almost invariably try far fewer methods than in prospective tests, resulting in much less new insight. Prospective tests, in contrast, force researchers to worry about a multitude of potential situations rather than only those observed in a known benchmark dataset. Prospective application in actual discovery projects, while important, also does not provide the necessary impetus, partly because often, the predicted compounds are in fact never tested [6] or the experimental data necessary to assess the quality of the predictions is absent—for example, because binding affinities are not measured or no crystallography is available.

To accelerate progress in quantitative predictive physical modeling, we need a series of community blind prediction challenges focused on pushing the limits of predictive techniques, beginning from problems which are just barely tractable with today's methods and advancing to problems just past the frontier. These challenges should be designed to have precisely the necessary high quality experimental data, but also be prospective, predictive tests. While the Drug Design Data Resource (D3R [23], discussed further below) provides

an existing community blind challenge on protein-ligand binding, it focuses on using *existing* pharmaceutical datasets for blind challenges, and not on introducing new data in a carefully controlled manner in order to maximize the learning value to the community [23]. In other words, D3R serves well to assess where we are now – but we need a carefully designed effort that will help the field achieve our goals.

In our experience, physical modeling advances most rapidly when confronted by specific problems which must be modeled accurately, as revealed by carefully collected and curated data. Thus, we need an effort which focuses on specific *component* problems of the overall problem of interest, collects and curates data that highlights these problems, and then brings this data to the community to drive progress via prospective challenges. This process allows the entire community to learn from what succeeds and fails in these challenges. The model we propose here has *already* worked to drive dramatic improvements in modeling, as evidenced by our **Statistical Assessment of Modeling of Proteins and Ligands (SAMPL)** series of challenges. SAMPL was born out of frustration with the lack of venues for comparing predictive accuracy on a level playing field and was initiated by Anthony Nicholls of OpenEye software in 2007/2008 [24], and has run approximately every two years since then [25–32]. Governance transitioned to an unfunded academic collaboration during SAMPL3 in 2012; this collaboration ran subsequent challenges as SAMPL4 (2014) and SAMPL5 (2016). The PI of this proposal (Mobley) played a key organizational role in SAMPL3–SAMPL5. SAMPL geared around the idea that learning collectively from failures and successes, rather than competition, will provide the greatest long-term rewards in terms of progress in the field. SAMPL has always involved a component focused on calculation of relatively straightforward physical properties such as hydration free energies, but also introduced host-guest systems as model binding systems for SAMPL3-5, supplementing protein-ligand challenges (on trypsin and HIV integrase) which appeared in SAMPL3-4. SAMPL has already been a tremendous resource for the community, resulting in roughly 100 publications (some are coming out as of this writing) which are typically cited 5-50 times or more each [refs].

Here, we expand SAMPL by **designing a new series of SAMPL challenges specifically to maximize learning value to the community**. Until now, this has been impossible, because SAMPL has been entirely unfunded; its very existence has required "donation" of data and time from various sources, hindering the ability to collect datasets tailored to our purpose. The incarnation of SAMPL proposed here will deliberately bridge the gap between calculations of simple physical properties that isolate forcefield inaccuracies from sampling challenges, like hydration—which can be calculated fairly accurately with today's methods [29]—and the D3R Grand Challenges on protein-ligand binding, which are a major source of consternation for the community so far [23, 33–35]. Unless this gap is bridged, there is the very real possibility that modeling may simply continue to fall far short of expectations in pharmaceutical challenges like D3R for reasons which are unclear. The extension of SAMPL proposed here will allow us to form blind challenges designed to highlight major reasons for failure and drive progress towards resolving them.

Our major goal is to rapidly advance predictive modeling to where it can guide experimental work doing biomolecular design, and extension of the SAMPL challenges will do exactly that. This work will play a vital role in enhancing the work being done on *existing* data by D3R, helping prepare methods for application to the more challenging systems emerging from pharma in D3R's challenges.

INNOVATION

Blind predictive challenges—and SAMPL in particular—have already led to important new science in the form of method development, evaluation, robustness, and force field development. But they have not yet resulted in a dramatic improvement in predictive molecular design, in part because the **lack of funding has left us unable to design a systematic and progressive series of challenges that drives such improvements**. Previous SAMPL challenges have been valuable in isolation – but they have not moved the frontier because they were driven by data availability rather than purposeful design. Here, we propose the innovative step of designing SAMPL to advance physical modeling.

Several historical examples serve to highlight how SAMPL can foster innovation (though far more examples are available [CITE]). The first several SAMPL challenges on hydration free energies had rather hit-and-miss performance, highlighting pitfalls of existing methods and force fields which led to marked improvements in PB models [refs], recognition of some limitations of fixed-charge force fields [refs], repair of some of these force field deficiencies via additional polarization or introduction of off-site charges [refs], and helped motivate alternate implicit or hybrid solvent models [ref]. Together these advances led to a marked improvement in accuracy for calculations of hydration free energies between SAMPLX and SAMPLY (Figure []). Shifts in protonation state and tautomer proved particularly important in the recent SAMPL5 logD challenge [refs]. This challenge provided a tractable opportunity to isolate and explore these specific physical effects, which are so important in protein-ligand binding, while avoiding the full complexity of pharmaceutical binding studies. Host-guest binding studies have also

been particularly important [ref benchmark sets], highlighting the importance of salt effects [refs] and in some cases revealing more severe force field limitations than observed in hydration and distribution challenges [ref Gilson group stuff], pointing the way forward for improving predictive models of molecular interactions [ref].

This work is also innovative because of the uniqueness of SAMPL. While there are other predictive challenges in the area of biomolecular modeling, such as D3R [ref], the pKa cooperative [ref], CAPRI [ref] and CASP [ref], **no other effort focuses specifically on driving quantitatively predictive protein-ligand modeling**. The SAMPL expansion we propose here is unique because it will be *specifically designed* to drive improvements in modeling accuracy for biomolecular design, rather than simply serving to evaluate accuracy on targets of pharmaceutical interest. It also plays an enabling role for a wide variety of other science, rather than functioning as a stand-alone entity. SAMPL benefits the whole modeling community—for example, protein-ligand docking software has improved as a direct result of SAMPL hydration challenges [ref Coleman], and commercial software vendors have introduced new features or scientific improvements based on participation in SAMPL challenges [cite Klamt stuff]. In effect, **SAMPL serves as an engine to spur innovation by new soliciting novel approaches to complex problems from the community and evaluating their success on blinded data.** The proposed project will ensure that SAMPL not only continues but becomes even more valuable to the community.

This work also focuses on innovative experimental methods. Specifically, in Aim 3, we develop a new informatics platform to facilitate the rapid identification and development of useful protein-ligand model systems that are both experimentally tractable for high-throughput biophysical measurements and focus on specific challenges of interest. We employ a fully automated wetlab to screen potential model systems for expression, carry out high-accuracy biophysical measurements, and perform automated error analysis to carefully assess experimental uncertainty. The Chodera lab—responsible for Aim 3—has substantial expertise in the area of automation of biophysical measurements and uncertainty analysis [36], and releases data analysis tools and 3D printed parts developed to facilitate these experiments as open source. **This work is at the forefront of innovation in high-throughput, automated biophysical experiments to produce high quality data with well-characterized uncertainties.** Not only will the data itself be of prime importance to the community, but the techniques themselves will help future experiments.

In Aim 4, we will not only run annual iterations of SAMPL community challenges, but also perform our own reference calculations with the latest techniques, testing their accuracy and using these to assess the current state-of-the-art. Both the Mobley and Chodera labs are experts in development of free energy methods for application to physical properties (e.g., [a couple refs]) and binding (e.g., [a couple refs]), and the **reference calculations we perform in Aim 4 are particularly important for innovation as well**, serving several key roles: (1) Benchmarking our latest method developments against current "best practices" methods (by doing calculations via both approaches); (2) Facilitating learning, allowing others to experiment with how a change in method or force field impacts results; (3) Focusing the field on key issues by doing sensitivity analysis to whether conditions such as ionic strength, protonation state, tautomer choice, etc., impact computed values.

Innovation in Aim 4 extends beyond reference calculations to analysis of the challenge itself. When methods differ in performance, it is critical to understand whether the differences are statistically significant and important, and to provide an accurate accounting of the uncertainty in performance measures. Thus, careful and innovative analysis of challenge outcomes is particularly important in SAMPL [refs], in some cases driving experimentation with new performance metrics [refs]. Tracking down *why* models are failing and what particular problems still need attention is an in important and powerful aspect of SAMPL. Much of this is left to the participants, but as organizers we do a great deal of work spotting patterns and providing a venue for participants to explore these issues. In Aim 4, here, we also extend the scope of our reference calculations, providing further opportunity to drive innovation in this way. Additionally, we try to draw attention to and promote analysis of model uncertainty (as distinct from statistical uncertainty) in calculated values [refs], as understanding the confidence levels of predictions is particularly important for guiding molecular design.

APPROACH

Our approach to systematically advancing modeling for biomolecular design involves collecting carefully targeted experimental datasets for challenges focusing on physical property prediction, host-guest binding, and protein-ligand binding that isolate individual limitations in quantitative physical modeling to solicit and evaluate multiple solutions from the community. Aims 1–3 focus on tailoring and generating this experimental data, while Aim 4 focuses on fielding annual SAMPL challenges. Aims 1–3 bring together multiple laboratories and both theorists and experimentalists: graduate students from the Mobley and Chodera laboratories are paired with well-equipped experimental groups in industry to collect physical property data (Aim 1); leading experimentalists in supramolecular chemistry Gibb and Isaacs work with theorist Mobley to perform host-guest affinity measurements (Aim 2); new

approaches to the automated development and investigation of model protein-ligand systems are developed in the Chodera lab (Aim 3). The annual SAMPL challenges organized by the Mobley and Chodera labs (Aim 4) will evaluate best-practices reference calculations, will feature both annual physical meetings (coordinated with D3R) and virtual communication platforms to maximize community engagement, develop novel statistical analyses to identify and respond to problem areas, release benchmark datasets alongside primary experimental data, organize special journal issues to report progress, and encourage tool automation whenever possible.

Aim 1: Collect new physical property datasets to assess accuracy and spur improvements in force fields and modeling of protonation states and tautomers.

Simple physical properties such as solvation, partitioning, and protonation equilibria can be calculated quite precisely with physical methods, allowing quantitative comparison of calculated values with experimental measurements and revealing and isolating deficiencies in our current physical modeling approaches. For example, these properties allow us to directly probe force field accuracy and chemical effects like protonation and tautomer handling in the absence of slow conformational changes and other affects which make this assessment so much more difficult in protein-ligand systems.

Rationale: We will generate new solution-phase physical property measurements for drug-like molecules to motivate improvements in force fields and handling of protonation states and tautomers. This builds on our work on water-cyclohexane distribution coefficients for SAMPL5, done in partnership with Genentech, which revealed major problems with handling of protonation states and tautomers [31] as well as serious forcefield limitations [37]. Our goal is to assess the current state of physical calculations of these properties, and drive innovation. Distribution coefficients serve well for this, as they relate to transfer free energies from aqueous high-dielectric to low-dielectric environments, capturing many of the characteristics of transfer of drugs from water into binding sites but absent challenges with receptor conformational sampling and specific ligand-receptor interactions. In SAMPL5, distribution coefficients were challenging enough that many methods performed poorly, with even the best methods having accuracies less than would have been expected based on hydration free energies in water [31], yet failures were informative and the major sources of error were issues which will also plague prediction of ligand-receptor interactions. In some respects, distribution coefficients posed the ideal SAMPL challenge, hitting the sweet spot in terms of difficulty—difficult enough that clear failures were frequent and there is much room for improvement, but not so difficult that the reasons for failure were unclear in general. Still, many models disagreed with experiment in consistent ways for particular compounds [31, 37–39], revealing the impact that targeted follow-up experiments (such as those we will conduct here) could have on improving models. Our choice of cyclohexane-water distribution coefficients rather than octanol-water was motivated by both complexities with octanol experiments ([40], making it difficult to ensure the system modeled is the same one simulated) and potential sampling problems in octanol [41], though challenges we propose here will broaden past cyclohexane.

This new experimental data is critical to maximize impact on the modeling community. While the community has already benefitted considerably from SAMPL, as indicated by the nearly 100 publications on SAMPL [refs], the citation count, and lessons learned, the benefits are not what they could have been if the effort had funding. A funded, coordinated effort allows the targeted collection of datasets designed to focus on the most important problems. With multiple investigators and collaborators, we are poised to respond and adapt to new challenges and opportunities which emerge in a manner which has not so far been possible for SAMPL. Additionally, we will be able to continue experimental work until the necessary data is collected rather than terminating it at a specific time point dictated by industry funding – allowing us to do things like ensure the full dynamic range of $\log D$ values is covered, unlike in SAMPL5 [31, 39]. This will improve our ability to learn from the data – for example, the lack of dynamic range for SAMPL5 meant that, when calculated values often spanned a larger dynamic range than the experimental values, it was unclear if this is an artifact of the limited range in the data set, or because of issues with the calculations or with the experiments themselves [31, 37–39]. With funding to extend the experimental set and the ability to do follow-up experiments (as we will have here) this would have been resolved, providing further impetus to improve models.

Below, we summarize tentative plans for SAMPL6-10, though we will adapt to new experimental opportunities as they arise and respond to persistently difficult challenges by ensuring the challenge is repeated with new data until clear improvement is seen.

SAMPL6: Cyclohexane/water and octanol/water distribution coefficients. Building on the success of distribution coefficient measurements in SAMPL5 and their surprising ability to motivate rapid advances in physical modeling methodologies [refs], we will measure cyclohexane-water distribution coefficients at pH 7.4 for a new batch of commercially-available drug- and fragment-like molecules for which no data is currently available. Given the routine nature of octanol-water distribution coefficient measurements and indications that their prediction may

be computationally tractable [40, 42] despite the heterogeneous structure of the wet octanol phase [41], we plan to also collect octanol-water distribution coefficient measurements for the same compounds. Because pKa prediction was difficult but critical for SAMPL5, we will measure and provide pKa values for the SAMPL6 compounds to focus on forcefield issues and tautomers, revisiting pKa prediction in SAMPL8.

SAMPL7: pKa measurements for drug-like molecules. While much less complex than protein-ligand affinities, distribution coefficient measurements still conflate several issues which are still complex, namely protonation state and tautomer prediction, as well as transfer into different environments which may contain small but important quantities of cosolvents. Thus, we will likely need to turn to separating these issues to improve our handling of them one at a time. For SAMPL7, then, our tentative plan is to measure pKa values for an extensive set of drug-like molecules in water and provide data specifically on pKa values, thereby separating the issues of predicting protonation state from those of transfer and paving the way for SAMPL8.

SAMPL8: pKa measurements and distribution coefficients. In the next challenge, we will re-combine the pKa and transfer issues in a way to maximize learning – specifically, we will not only measure distribution coefficients but also measure pKa values for the same set of compounds, allowing participants to (a) predict distribution; (b) predict pKa; and (c) predict partitioning.

SAMPL9 and SAMPL10. We will focus on solubility prediction in SAMPL9, and membrane permeability in SAMPL10, because of new computational developments. With solubility predictions now becoming tractable [9–11] (with Schrödinger also working on amorphous solubility prediction), solubility measurements will be a valuable test for SAMPL9, combining the solvation aspects of SAMPL1-8 with a new solid phase component. New computational techniques are targeting membrane permeability [13, 14], and this is experimentally accessible (see support letters from Pfizer and Merck), leading to our interest in permeability for SAMPL10.

Experimental plan: Experimental data will be collected in collaboration with partners in the pharmaceutical industry, roughly following the model used for SAMPL5, where the Chodera lab sent a graduate student (Bas Rustenburg) to Genentech to conduct cyclohexane-water logD measurements by adapting a high-throughput mass spectrometry workflow already in use within Genentech [39]. To collect physical property measurement datasets for SAMPL6-10, the Mobley and Chodera labs will send graduate students on several-week visits or internships to industry collaborators to collect targeted datasets. Working with industry collaborators (see Letters of Collaboration from Genentech, Pfizer, and Merck) gives us substantial access to equipment and high-throughput measurement workflows—such as the Sirius T3 from Sirius Analytical (which can measure partition/distribution coefficients, pK_as, and solubilities for molecules with titratable groups), automation equipment, and compound libraries—for the purposes of rapidly collecting targeted datasets. As noted, we already know that this model will work given our experience in SAMPL5 [39], and our pharma partners see the value of this data and the SAMPL challenge to the modeling community.

Overall, Aim 1 extends prior SAMPL challenges via data focused on quantitative prediction of physical properties of tremendous relevance to accurately predicting biomolecular interactions, paving the way to applications in more complex systems addressed in Aims 2 and 3.

Aim 2: Measure affinities of drug-like compounds in supramolecular hosts to challenge quantitative models of binding in systems lacking major receptor sampling issues.

The physical property challenges of Aim 1 focus on the behavior of small molecules and how this is modulated by changes in environment, in the absence of receptors and the associated potential for slow sampling, strong specific interactions, and other challenges such as salt effects. Binding in host-guest systems retains many of the same challenges seen in Aim 1 and introduces strong specific interactions and other challenges like salt effects [43], while still avoiding many of the issues with slow sampling (of protein conformational changes, ions, and ligand binding modes) seen in protein-ligand interactions. Thus, new data for SAMPL challenges on host-guest binding is critical to provide challenges of intermediate complexity between those of physical properties and those in bioolecular binding. —but binding in host-guest systems introduces a wider variety of challenges relevant to biomolecular interactions, but without the full array of challenges seen in protein-ligand interactions, as reviewed recently Thus, we believe that new data and SAMPL challenges on host-guest binding are a critical step towards accurately modeling biomolecular interactions, focusing the field on issues not commonly en-

drug	features
memantine	adamantane; 1:1
saxagliptin	adamantane; 1:1
premarin	steroid
pancuronium	steroid
varenicline	1:1 vs 1:2
valsartan	pKa 4.37
omeprazole	pKa 4.77
ranolazine	pKa 7.17; epitopes
pradaxa	pKa 3.87; epitopes
nilotinib	epitopes; pKa 6.3
sensipar	epitopes; folding
vyvance	diamine; epitopes; folding
minocycline	tetracyclin; amino aniline

Table 1. **Selected drugs whose binding to CB[n] hosts will be assayed for SAMPL6, 8, and 10 challenges (SA 2.1).** These drugs bind to the cucubituril-based host systems considered here, some at high affinity, so measuring their affinities provides a way to test methods for predicting binding interactions absent complexities present in protein-ligand systems.

countered in physical property challenges (such as the importance of accurately modeling ionic conditions) that are highly relevant for protein-ligand interactions but easily overlooked. Already, over the past several SAMPL challenges, host-guest systems have provided key tests for modeling of binding interactions, resulting in new attention paid to how co-solvents and ions modulate binding (resulting in errors of up to 5 kcal/mol when these effects are neglected) and the importance of adequately sampling water rearrangements [30, 32, 43, 44]. Host-guest binding proves remarkably difficult to model accurately [45], in part due to force field limitations (resulting in new force field work [46]).

Here, we design a series of SAMPL challenges focused on two general classes of host-guest systems—cucurbiturils and derivatives (SA 2.1) and Gibb's deep-cavity cavitands (GDCCs, SA 2.2)—both of which build on studies from prior SAMPL challenges. These two sets of systems exhibit different challenges as recently reviewed [43], with the hosts of 2.1 bringing relatively modest co-solvent and ion effects but some receptor sampling problems for the acyclic hosts, and the GDCCs of 2.2 bringing profound ion and co-solvent effects as well as water sampling challenges. Methods which perform well on one class may not perform well on the other [43], since the distinct sets of challenges highlight different limitations. This diversity will drive innovations in far more areas than if we focused on a single host class.

Subaim 2.1: Cucurbituril-based receptors as model binding systems

Cucurbituril derivatives for host-guest binding. Building on previous success with cucurbit[n]uril (abbreviated CB[n]) experiments for previous SAMPL challenges [47–49], we will conduct a series of new experiments on these receptors for five new SAMPL challenges, with experimental work conducted by co-investigator Isaacs, an expert on these systems who has provided data for previous SAMPL challenges. CB[n] receptors are particularly well suited for the SAMPL challenges because they exhibit: (1) high binding affinities toward suitable guests in water comparable to protein-ligand affinities (routinely μM to nM; occasionally pM to fM) [50–56], (2) high selectivities between structurally related guests which translate into large $\Delta\Delta G$ values [57], (3) low molecular weights (1–2 kDa) permitting high levels of theory to be used, and (4) highly restricted conformational degrees of freedom, eliminating challenges associated with protein targets with many soft degrees of freedom and slow motions. For SAMPL6-10, we will resynthesize a series of CB[n]-type receptors of increasing complexity, measure K_a values, and determine host-guest stoichiometry and geometry toward pharmaceutically relevant guests (selected drugs) in order to stringently test methods for predicting binding. Figure 1 shows the chemical structures of three hosts—Me$_4$CB[8] [58], glycoluril hexamer [59], and acyclic CB[n]-type receptors [60–65] which span the range from preorganized macrocyclic host to uncharged acyclic but preorganized host to highly charged acyclic host.
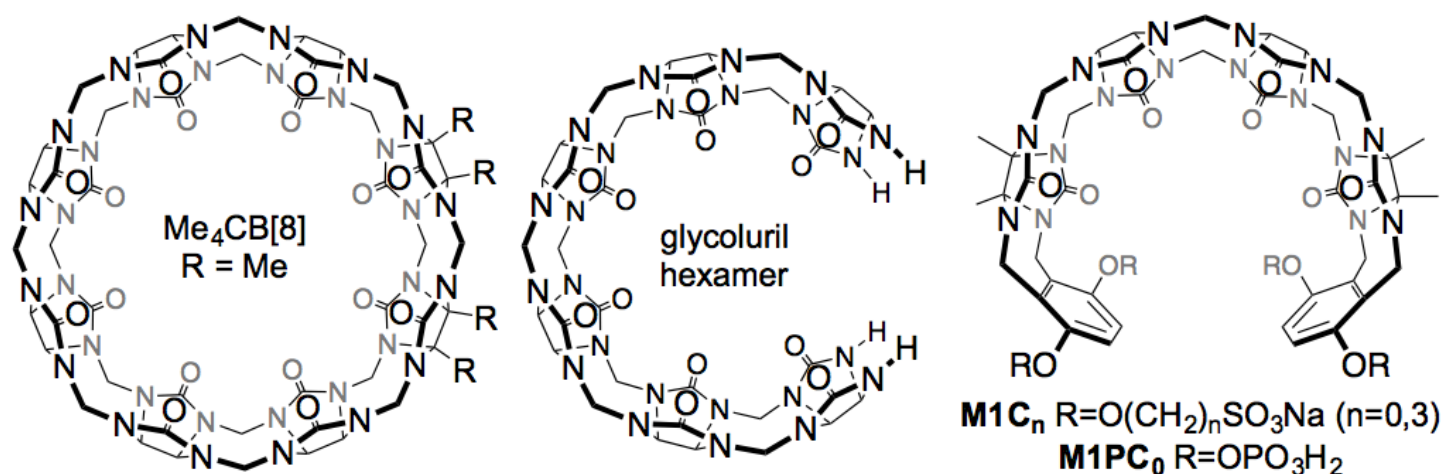

Figure 1. Structures of Me$_4$CB[8], glycouril hexamer, and acyclic CB[n]-type receptors. These receptors bind a variety of drug-like molecules, some with high affinity, and this binding will form the basis of a cucubituril-based series of SAMPL binding challenges.

SAMPL6-10 cucurbituril challenges. *For SAMPL6*, we will measure K_a and ΔH values, stoichiometry, and geometry for the interaction of Me$_4$CB[8] (a soluble CB[8] derivative) with 15 guests (selected top drugs, Table 1) by either direct or competition isothermal titration calorimetry (ITC), UV/Vis or fluorescence indicator displacement assay, or NMR competition experiments, as previously [49–51, 66]. Our selection of Me$_4$CB[8] binding to top drugs allows us to modulate the computational complexity by: 1) changing host flexibility (e.g. Me$_4$CB[8] can exhibit ellipsoidal deformation) [58], 2) allowing the possibility of binary or ternary (e.g. 1:1 and/or 1:2 host:guest)

complexes [67–69], 3) using drugs with several potential binding epitopes to induce sampling issues. Host:guest stoichiometry and geometry (e.g., which binding epitope is complexed) will be addressed by ITC n values, Job plots monitored by UV/Vis or NMR [70], and by ^{1}H NMR complexation induced changes in chemical shifts [71]. All three sets of studies will be conducted in phosphate buffered saline (pH 7.4 with physiological salt) which introduces its own complexities due to salt competition for binding [43, 72]. *SAMPL7* will revisit the same host, but use 15 different guests be selected from commercial sources on the basis of reference calculations (on a larger set of guests) to ensure that they cover substantial dynamic range and/or exhibit affinities that depend substantially on the force field or water model, thus effectively testing our force fields and methods. For *SAMPL8*, we will focus on binding of the same 15 drugs (Table 1), but to glycoluril hexamer. This is an ideal next step as this host exhibits increased conformational dynamics, and influences the number and energy of solvating (and unusually coordinated) water molecules implicated in the high binding constants for CB[n]-guest complexes [56, 73]. The selected drugs include several with pK_a values in the 3.8 to 7.4 range; given that CB[n]-type receptors (like biomolecular receptors) can induce pKa shifts in their guests of up to 4 pKa units [74–76], this will test how well models can predict these effects. Additionally, it will couple nicely with the focus on pK_a values in Aim 1. *SAMPL9* will revisit glycouril hexamer with the same 15 guests from *SAMPL7*. *SAMPL10* will shift to acyclic CB[n]-type receptors (e.g. M1C$_3$, M1C$_0$, and M1PC$_0$ that contain anionic solubilizing groups attached via different linker lengths. As in SAMPL3 [77], these acyclic CB[n]-type receptors introduce conformational complexity, and water interactions play a key role. Moreover, the presence of 4 anionic groups near the cavity will likely impact balance between ion-dipole interactions and the solvation of the free host.

Subaim 2.2. Gibb deep cavity cavitands for host-guest studies

History of octa-acid SAMPL challenges. During SAMPL4 [78] and SAMPL5 [79] we focused on two hosts: the octa-acid 1 (R = H) and another octa-acid derivative with four methyl groups positioned at the portal of the binding pocket (1, R = Me). These studies used isothermal titration calorimetry (ITC) to measure the thermodynamics of (1) host **1** (R = H) complexing a range of 9 carboxylate guests, and (2) the binding of 6 carboxylate and trimethylammonium guests to both hosts (**1**, H = H and Me; Figure 2). In both cases ^{1}H-NMR titration was also used to confirm ITC-derived free energies of binding. As noted above, co-solvent effects and water rearrangements posed particular challenges for predicting binding in these hosts. SAMPL5 emphasized how differences in the shape of the hydrophobic pocket of the host can have a profound affect on affinity for some guests [80].
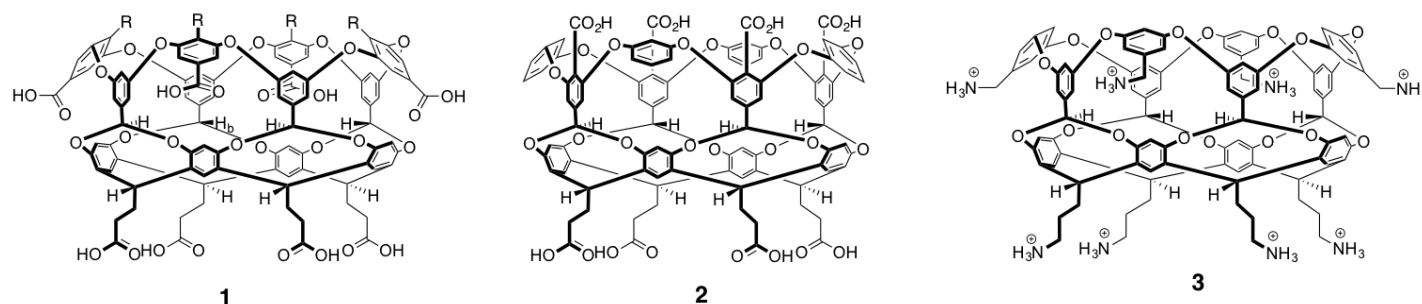

Figure 2. **Gibb deep cavity cavitands for SAMPL6-10.** These hosts bind a variety of carboxylate and trimethylammonium guests in a strongly salt-dependent manner, providing a stringent test of our ability to model binding and how it depends on these effects.

Novel deep cavity hosts probe the effects of binding site charge constellations. For future SAMPL challenges, we will expand on the range of hosts by including **2** and **3** in our ITC studies (Figure 2). Like cavitand **1**, host **2** is an octa-acid derivative. However, the four benzoate groups are relocated from the extreme exterior in the case of **1**, to the rim of the binding pocket in **2**. We surmise that this will have a direct effect on the binding of charged guests, but more subtly, an indirect effect on guest complexation via changes to the solvation of the empty host. Octa-trimethylammonuim cavitand ("positand" **3**) has the same overall architecture as host **1**, but inverts the charges on the water solubilizing exterior coat. While it is not yet clear if this switch in groups relatively remote from the pocket will directly affect guest complexation, results from related systems suggest it can (unpublished). Guests will be selected from commercial sources on the basis of reference calculations (on a larger set of guests) to ensure that they cover substantial dynamic range, marked differences in affinities between hosts, and/or affinities that depend substantially on the force field or water model, thus effectively testing our force fields and methods.

SAMPL6-10 deep cavity cavitand challenges. The host-guest challenge for SAMPL6 will focus on how well the effect of host carboxylate substituent location can be predicted, and will involve hosts **1** and **2** with a set of five, previously uninvestigated guests. Guests will be selected from commercial sources on the basis of reference

calculations in a similar manner to SAMPL7 in Subaim 2.1, specifically picking guests which have broad dynamic range and, here, have marked differences in affinities between hosts. SAMPL7 will provide a second iteration of this experiment to test algorithmic improvements in predictive modeling following SAMPL6 by comparing hosts **1** and **3** with a different set of guests. We anticipate that because of the relative remoteness of the charged groups in these two hosts, the effects of switching charges will be subtler than the differences between **1** and **2**. SAMPL8 will consider the effect of common biologically-relevant counterions/salts salts on guest binding, comparing the effects of NaCl and NaI on the complexation of five guests to **1**. We have previously shown that iodide has a weak affinity for the binding pocket of **1**, whilst sodium ions have an affinity for the outer carboxylates [81], requiring modeling to capture the differential affinities of these ions in addition to guest affinities to successfully model the observed affinities. SAMPL9 will follow up on this by examining the effects of these same two salts on the complexation of five guests to **3**, again giving the modeling community time to incorporate algorithmic improvements following SAMPL8. While we have not yet quantified salt affinities to host **3**, we expect the iodide to have affinity for both the pocket and the positively charged solubilizing groups. For SAMPL10 we will consider the effects of co-solvents on the binding of five guests to **1** and **2** to probe the effect of co-solvent competition for the binding site, as well as effects cosolvents may have in weakening the hydrophobic effect. While the number of guests considered in each challenge is relatively small, the total number of binding affinities measured is significant across the full family of hosts, meaning that the full data set will be of considerable value as a benchmark set [43].

Aim 3: Develop model protein-ligand systems that isolate specific modeling challenges of drug targets. A major goal of our effort is to drive advances in the quantitative modeling of protein-ligand interactions. While D3R [23] provides community blind challenges for biomolecular targets of pharmaceutical interest, these targets generally exhibit a daunting number of complexities that frustrate the ability for current methodologies to achieve quantitative accuracy, resulting in poor performance [23, 33–35]. For example, while kinases are targets of great interest to drug discovery, blind challenges involving kinase targets conflate issues of slow protein conformational dynamics [82], protonation state effects of both protein [83] and ligand [84, 85], charged ligands, and the modeling of complex divalent salt environments and phosphorylation state effects along with the standard computational challenges of conformational sampling and forcefield accuracy. Thus D3R exercises serve the community well as a benchmark of current accuracy, but **the ability of blind challenges on complex pharmaceutical targets to rapidly advance quantitative predictive modeling is limited.**

[DLM: The problem I think isn't just that we do badly on protein-ligand systems, but that we do badly and don't understand why. The paragraph above seems to indicate that the problem is just that these systems are complex and we do badly, but we do badly on host-guest systems too. The issue here is that the systems are so complex we don't know WHY we do badly. Can we highlight that?]

Instead, **we will identify model protein-ligand systems which *isolate* individual accuracy-limiting effects in successive cycles of prospective blind community challenges.** [DLM: See comment below about lysozyme – I think we may need to give our working definition of "model" in the first sentence or two here. One thing I've gotten from people (including Marti and Peter Preusch) in discussing the idea is that people are not going to be that excited about developing more "model" systems in the sense that lysozyme is a model system – that is to say, systems that are completely artificial. I think we may need to hit this objection early on by saying exactly what we mean by model – here, I think it's that we're talking about biological systems (and simple variants thereof) that closely resemble pharmaceutical targets but which we understood better.] This process focuses the field on identifying and evaluating multiple solutions to the accuracy-limiting effects (such as how to deal with ligand and protein protonation-state issues [90], slow protein conformational dynamics, etc.) free from other complicating factors, allowing a direct evaluation of how well the phenomena of interest are modeled. Datasets collected for these blind challenges then become standard benchmark datasets for retrospectively examining the effectiveness of modeling approaches in treating these effects to facilitate comparisons of methodologies in publications, while future iterations or variations of the same SAMPL experiment allow iterative refinement and prospective blinded evaluations of methodologies. In this way, this cycle of blind challenges utilizing model systems can rapidly drive progress in rapidly overcoming scientific hurdles limiting quantitative accuracy.

[DLM: These last two sentences seem to be general to SAMPL. Aren't they covered by Aim 4? If not, propose an edit there to incorporate these ideas?]

While model protein-ligand systems (such as the Shoichet T4 lysozyme mutants [43, 91]) have had ample success in driving progress in individual research laboratories, their power in blind community challenge cycles is amplified by leveraging community participation. [DLM: A lot of pharma folks laugh when lysozyme is mentioned, in part because they see it as trivial/ridiculous. I worry that if you use lysozyme to motivate this, people are going to connect the word "model" with lysozyme and think we are proposing creating more systems like lysozyme –
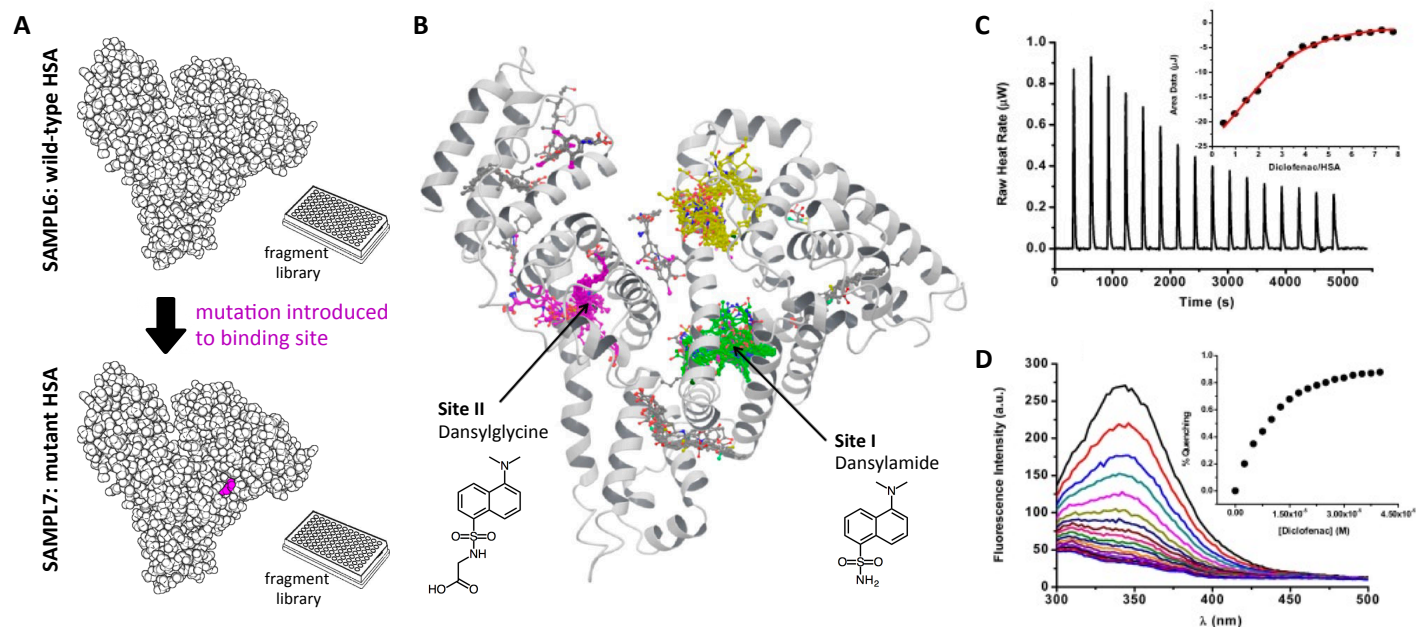
Figure 3. **The SAMPL6/7 protein-ligand challenge focuses on soluble drug fragment binding to human serum albumin (HSA).**
(A) SAMPL6 will introduce recombinant human serum albumin (HSA) as a target, against which a library of 96 small soluble drug fragments will be assayed. By introducing a mutation in one of the binding pockets, we will create a second challenge target for SAMPL7. *(B)* HSA, the most abundant plasma protein, has at least eight known binding sites, with major well-characterized sites (green, Sudlow's Site I; purple, Sudlow's Site II) which bind a variety of drugs, often modulating their pharmacokinetics [86] (figure from [87]). Two fluorescent probes—dansylamide and dansylglycine—have been shown to bind with $\sim\mu$M affinity and high selectivity to Site I and Site II, respectively; both exhibit binding-enhanced fluorescence at 480 nm, and can be used to site-specifically probe ligand affinities by competition. *(C)* Binding affinities of soluble molecules can easily be measured by isothermal titration calorimetry (ITC); here, the ITC titration of HSA by diclofenac (a Site II ligand [88]) is shown [89]. *(D)* HSA tryptophan fluorescence quenching can also be used to measure ligand binding affinity; here, HSA titration by diclofenac is shown, with the inset plot showing percent quenching at 346 nm [89]

not biologically relevant, far too easy, etc. See note above. The solution may be to just explain what we mean by "model" very early in this section and deal with the objection head-on.] For example, SAMPL3 featured the binding of small, rigid charged molecules to bovine trypsin [92], and rapidly focused the field on the deficiencies of current alchemical free energy methodologies in treating the binding of charged ligands. Within two years, multiple laboratories had developed and disseminated convergent practical solutions to effectively handle charged ligand binding [**?**, 93, 94].

SAMPL6-10 model protein-ligand challenges. For the SAMPL6-10 challenges, we propose to introduce a new model protein-ligand system each year, with challenges fielded for each system for at least two consecutive years to allow iterative methodology improvement and assessment. Immediately following the challenge, the data (including all primary data) will be published and released as a version-controlled benchmark dataset for retrospective evaluation. The first challenge (introduced in SAMPL6) will focus on modeling the binding of small soluble drug fragments to a relatively rigid protein with multiple weak binding sites, isolating the ability of current-generation modeling approaches to model weak and multiple binding effects. Because rapidly focusing the field on current challenges in predictive modeling requires the ability to adapt to deficiencies identified by D3R/SAMPL challenges of the previous year, subsequent model systems will be rapidly identified and developed using a new informatics platform we have developed to identify tractable model systems.

SAMPL6: Assessing predictive modeling of binding to multiple weak sites via measuring fragment binding to human serum albumin (HSA). Human serum albumin (HSA), the most abundant blood plasma protein, has the remarkable ability to bind a great variety of small molecule drugs in multiple binding sites (Figure 3B) [86]. As a result, HSA is not only an excellent model system for isolating the challenge of binding multiple weak ligands to a stable rigid protein, it is also a pharmacologically relevant system due to its ability to drastically modulate drug pharmacokinetics [87]. HSA has at least *eight* known binding sites, with numerous crystal structures available for drugs binding to two predominant sites (Sudlow Site I and II) [87]. Small soluble molecules resembling drug fragments have previously been shown to have a high likelihood of detectable binding to HSA ($\geq$90% of small druglike fragments, as detected by SPR [95]), providing an experimentally-tractable diverse set of ligands spanning several orders of magnitude in affinity [CITE]. [DLM: What citation here for "several orders of magnitude"?] As cur-

rent advanced methodologies such as alchemical free energy calculations currently assume a single well-defined binding site with high affinity [96], this dataset will allow the isolation of the effect of weak multiple binding from the majority of other counfounding factors in protein-ligand binding. However, HSA is relatively rigid, and computational methods show some promise in computing binding affinities to HSA [?], so this is not too difficult as a starting challenge. [DLM: Did you have other references we should cite here? I think we need to make the point that while there are complexities, there aren't TOO MANY...]

Recombinant HSA will be expressed in *E. coli* and purified via refolding from inclusion bodies [97], and will be defatted at low pH to ensure the resulting protein is free of the complications of both glycosylation and bound fatty acids found in plasma-isolated HSA [98]. Recombinant expression will also allow a mutant form of HSA (engineered via quick-change single-primer mutagenesis) to be fielded for a SAMPL7 iteration of this challenge (Figure 3B). We will obtain a diverse library of 96 novel soluble drug-fragment-like molecules in pre-plated format as dry compound, and assay it for HSA binding using automated isothermal titration calorimetry (ITC) (Figure 3C), with the goal of characterizing the overall binding affinity of the compound to HSA. The same ligands pre-plated in DMSO format will be used to conduct a separate set of fluorescence titration assays (monitoring tryptophan fluorescence quenching) and competition assays in which the site-specific fluorescent probes daynsylamide (Site I) and dansylglycine (Site II) will be used to measure site-specific affinities for Sites I and II (Figure 3D), allowing participants to validate whether they predicted the correct binding site and, if so, the site-specific affinity.
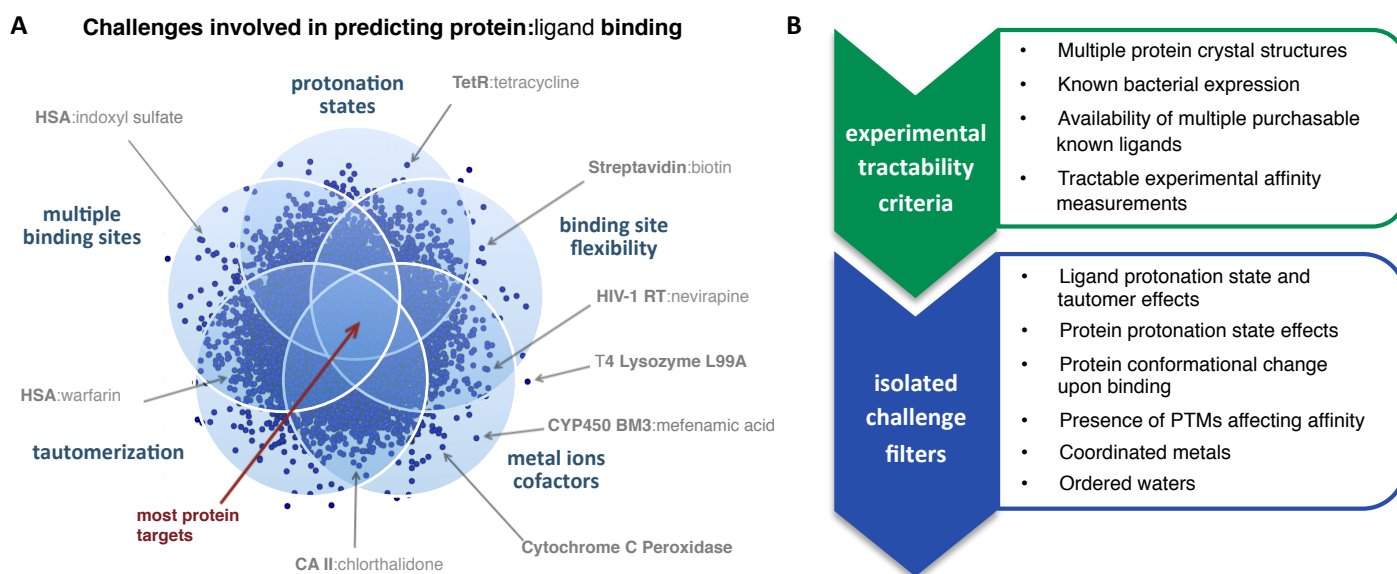

Figure 4. **Mining model protein-ligand systems to focus on individual modeling challenges via a structural and chemical informatics platform.** SAMPL7-10 will feature the introduction of new model protein-ligand systems designed to focus on individual challenges judged to be of critical immediate importance following current D3R/SAMPL blind competitions. *(A)* Since most protein targets of pharmaceutical interest feature a multitude of simultaneous challenges, our goal is to identify model protein targets that isolate individual effects to focus community efforts by fielding new blind challenges. Some example protein-ligand pairs and the conceptual challenge categories they fall under are shown in the figure: T4 Lysozyme L99A [21], HSA [99], Carbonic anhydrase II (CAII) [99,100], Cytochrome C Peroxidase [101], Cytochrome P450 BM3 M11 (CYP450 BM3) [102], HIV-1 Reverse Transcriptase (HIV-1 RT) [103,104], Streptavidin [103], Tet repressor protein (TetR) [103,105]. *(B)* In order to rapidly develop new experimentally and computationally-tractable model protein-ligand systems, we have developed a structural and chemical informatics system that applies successive filters to all potential protein-ligand systems for which structural data is available. Suitable model systems should meet all experimental tractability criteria (green box) and possess only a few challenging properties, ideally only one (blue box).

SAMPL7-10: Rapid, responsive development of new model systems using a novel informatics platform.
We have developed a novel informatics platform called TargetExplorer aimed at identifying new protein targets that can be rapidly developed into experimentally- and computationally-tractable model systems focusing on individual challenges. This tool—which will be made accessible to other laboratories via an easy-to-use web interface during the course of this project—successively filters all protein-ligand complexes identified in the PDB according to a list of criteria that allow facile development as a model protein-ligand binding system, as well as identification of systems that isolate individual challenges in modeling accuracy. [DLM: Rephrase long, cumbersome prior sentence? I can't even figure out if it's grammatically correct – I keep thinking "as a model protein-ligand binding system" should be "as model protein-ligand binding systems" but I can't figure out what it modifies; I think it's "complexes" in which case it should be plural, "systems"? Often, the easiest solution to this type of thing is just to make your sentence simpler. Seems like it's three ideas that maybe should be three sentences: (a) this

tool will be made available via the web; (b) it filters all complexes; (c) the filtering is done to ensure they can be pursued as model systems and will be informative... Experimental tractability includes: (1) the availability of multiple protein-ligand crystal structures; (2) known bacterial expression (e.g. from PDB EXPRESSION_SYSTEM records); (3) the capacity to bind a wide dynamic range of ligands (determined via data available in ChEMBL); (4) the availability of multiple known ligands that can be purchased (via ZINC); (5) tractability of experimental affinity measurements, such as known ligands with potentially fluorescent scaffolds (for fluorescence competition assays), highly soluble ligands (for ITC), or ligands above a minimal mass (for SPR or MST). A number of additional filters annotate potential experimentally tractable systems for suitability as a model system that isolates individual challenges, such as: charged ligands or potential ligand protonation state or tautomer effects [99] (deduced from predicted aqueous protonation/tautomer energies); potential protein protonation state effects (deduced from MCCE2 calculations [106]); protein conformational changes (deduced from variation in protein conformation or the presence of unresolved loops in protein-ligand crystal structures); the presence of post-translational modifications that may affect affinity (deduced from Uniprot annotations); coordinated metals (identified in crystal structures); ordered waters (present in multiple crystal structures); etc.

The Chodera lab has developed an automated wetlab for the purpose of rapidly developing model protein systems using bacterial expression techniques (see Equipment and Facilities). Potential targets matching desired challenge criteria will be screened for bacterial expression using high-throughput cloning, transformation, and expression testing, with purity and yield assessed by capillary electrophoresis on a Caliper GXII. Targets will be screened for stability in various buffers using Thermofluor [107]. Ligands identified via the informatics platform as spaning a wide dynamic range of binding affinities will be purchased as dry powder stocks and prepared for assay by highly accurate gravimetric solution preparation techniques using a Quantos automated balance. A wide variety of biophysical techniques are available to provide accurate, quantitative measurements of protein-ligand binding affinities, including fluorescence (if fluorescent probe ligands are available), absorption (e.g. Soret band shifts), automated isothermal titration calorimetry (provided ligands are sufficiently soluble), surface plasmon resonance, microscale thermophoresis (MST), luminescence, and alphascreen; all except MST are fully automated. [DLM: "are available" is a case where passive voice is hurting you – I can't tell if you mean you have all of these techniques available in your lab, or if they just exist somewhere. I presume you mean in your lab, but you could just say "we have a variety of techniques available to..."

Our approach to developing challenge datasets will be twofold: First, small molecules similar to known ligands will be purchased and assayed, with the presumption that these molecules are likely to have measurable affinities. Second, site-directed mutants will be introduced to modulate the binding affinities of known ligands using single-primer quick-change mutagenesis, which can be performed and screened for expression in 96-well format. Challenge datasets will therefore consist of a matrix of protein mutants and ligands, providing a rich dataset to deeply explore the effects of interest.

Aim 4. Field iterated community blind challenges to advance quantitative biomolecular design. While Aims 1–3 focus on the critical need for targeted datasets to improve deficiencies in physical modeling of protein-ligand interactions, **the value of this data is amplified enormously in the strategic release of this data through multiple iterations of coordinated SAMPL blind challenges and related activities**. These blind challenges are designed to test the state of the art, provoke new methodological and force field innovations, allow comparative evaluation of methods, and drive downstream improvements. The new, progressive, targeted nature of the data generated for this purpose means that SAMPL challenges can now build on one another, and for success in later challenges, participants must build on lessons learned from prior challenges. SAMPL challenges and subsequent data release activities will therefore facilitate rapid iterations of application, learning, and improvement.

SAMPL blind challenges. SAMPL challenges will have submission deadlines yearly, though not every data type or component of Aims 1-3 will be annual. The full timeline for SAMPL challenges (Figure 5) will be made available on the SAMPL website [https://drugdesigndata.org/about/sampl] at the outset, allowing participants to plan their work and select what challenges to be involved in. As experimental data for each component becomes available and is curated, input files and challenge details will be made available online (and advertise via e-mail lists from prior SAMPLs and CCL), at least six months prior to the challenge deadline; data not yet available at that time will be held for a subsequent challenge (with the exception of three months for year 1 due to startup timescales). As in prior SAMPLs, submissions will be handled by an automated web upload service on the SAMPL website (which will be migrated to separate hosting if the D3R effort is not renewed after its current term) which validates submissions to ensure that they meet format standards we specify along with the challenge details. As in SAMPL4 and SAMPL5, analysis will also be conducted by our automated Python framework, and results will be returned automatically via a web framework. Participants will be allowed to submit multiple sets of predictions per SAMPL component to allow for comparative assessment of models. One new ingredient will be that each participant will

be required to designate one submission as their "primary" submission which will be included in the formal ranking of performance; other submissions will still be analyzed and receive an assessment of performance, but only one will be formally ranked. We find that participants learn a great deal from being able to compare multiple methods or protocols, but providing some participants with more "shots on goal" than others in the formal ranking can be seen by some as unfair. All participant submissions and methodology descriptions will (as before) be made available publicly on the website, along with participant information (except for participants who specifically request to remain anonymous prior to submission). Aggregate statistics and historical performance will also be made available on our website, along with a record of linking publications to historical submissions.

Our goal is not just to run blind challenges, but to advance modeling by helping participants identify both modeling failures and potential solutions. To achieve this, we provide guidance to participants as to what known modeling issues we expect may be relevant when providing details on each SAMPL component. For a host-guest system, for example, we might highlight known buffer/salt effects, protonation state challenges, and point out previous work on sampling challenges, with pointers to the relevant experimental work and to modeling work from past SAMPL challenges and elsewhere [43]. This helps participants design their approach. **Additionally, we will run reference calculations using current best practices.** This serves several purposes: It provides a test of the current methods we select (usually those we view as current standard best practices) and current force fields; it helps facilitate learning—we announce what calculations we plan to perform, make input files available in a wide variety of formats [?, 31, 32], and others can repeat our calculations with a different method but same system and force field to compare methods, or swap force field but keep the method and system fixed to compare force fields, etc.; and it allows us to conduct sensitivity analysis, as by varying the conditions of our simulations (protonation state, tautomer, etc., [31]) we can see how much this impacts calculated values and thereby how important it is, even if participants don't do these tests. Reference calculations have, for example, helped us highlight the importance of a small amount of water in cyclohexane for accurately calculating log D values, determine that an incorrect tautomer could affect calculated values by many log units [31], and highlight how forcefield modifications could significantly improve results on hydration free energies [29]. To further aid follow-up studies, we will make the input files, results, and simulation workflows used for our reference calculations—along with the data—available via open software development portals such as GitHub and Docker Hub.

Physical methods are only valuable if they can reliably outperform alternate methods, so **a new focus of SAMPL6-10 will be selecting quality null models and running them to provide a point of comparison for participants**. While SAMPL efforts in the past have used some null models, lack of manpower has required that these be particularly crude (such as "guess zero" models [refs]) which provide no predictive value. In this work, we will announce our selected null models when we open each challenge component, so participants know what their results will be compared with. Statistical analysis of SAMPL performance, and comparison with nulls, will continue to be an important component of our work [refs].

Following submission and analysis of each SAMPL challenge, challenge results will be released and discussed, with SAMPL workshops allowing more formal presentations on and discussion of results in years 1, 3, and 5. Workshops will run every two years at the request of past participants, and will be co-run and co-hosted with D3R Grand Challenge workshops (see support letter). During the off years, SAMPL challenges will still run, but discussion of and dissemination of results will be via asynchronous means (as discussed below) and a "virtual workshop" consisting of talks and interaction over Google Hangouts or YouTube Live (formerly Hangouts On Air). While coordination with D3R will mostly be at the level of workshops, we will also work with them ensure that SAMPL challenge submission deadlines are offset from deadlines for D3R to allow maximum community participation in both efforts.

Dissemination of results and data. Rapid dissemination of results is critical so that new insights can be rapidly spread to the community. We will continue to publish special issues of the *Journal of Computer Aided Molecular Design* (JCAMD) collecting publications related to each year's SAMPL challenges (see Letter of Support). To ensure immediate availability of reports, we will strongly encourage prepublication sharing of results and analysis, including both slides and posters from SAMPL meetings (via F1000 Research) and paper preprints (via bioRxiv). We also want to ensure that participants are able to learn from one another by exchanging ideas outside of formal workshops and meetings. While this has happened in the past—for example, when participants using similar methods work together after the SAMPL meeting to identify the origin of these discrepancies [refs]—we hope to accelerate this kind of collaboration. To facilitate more open communication between the community, we will use collaboration software—such as Slack, which facilitates scientific communication for the NASA/JPL Mars Rover teams and NSF antarctic scientific research teams—to build a community discussion platform, allowing interested individuals to openly exchange ideas, tests, and work together to learn from their work.
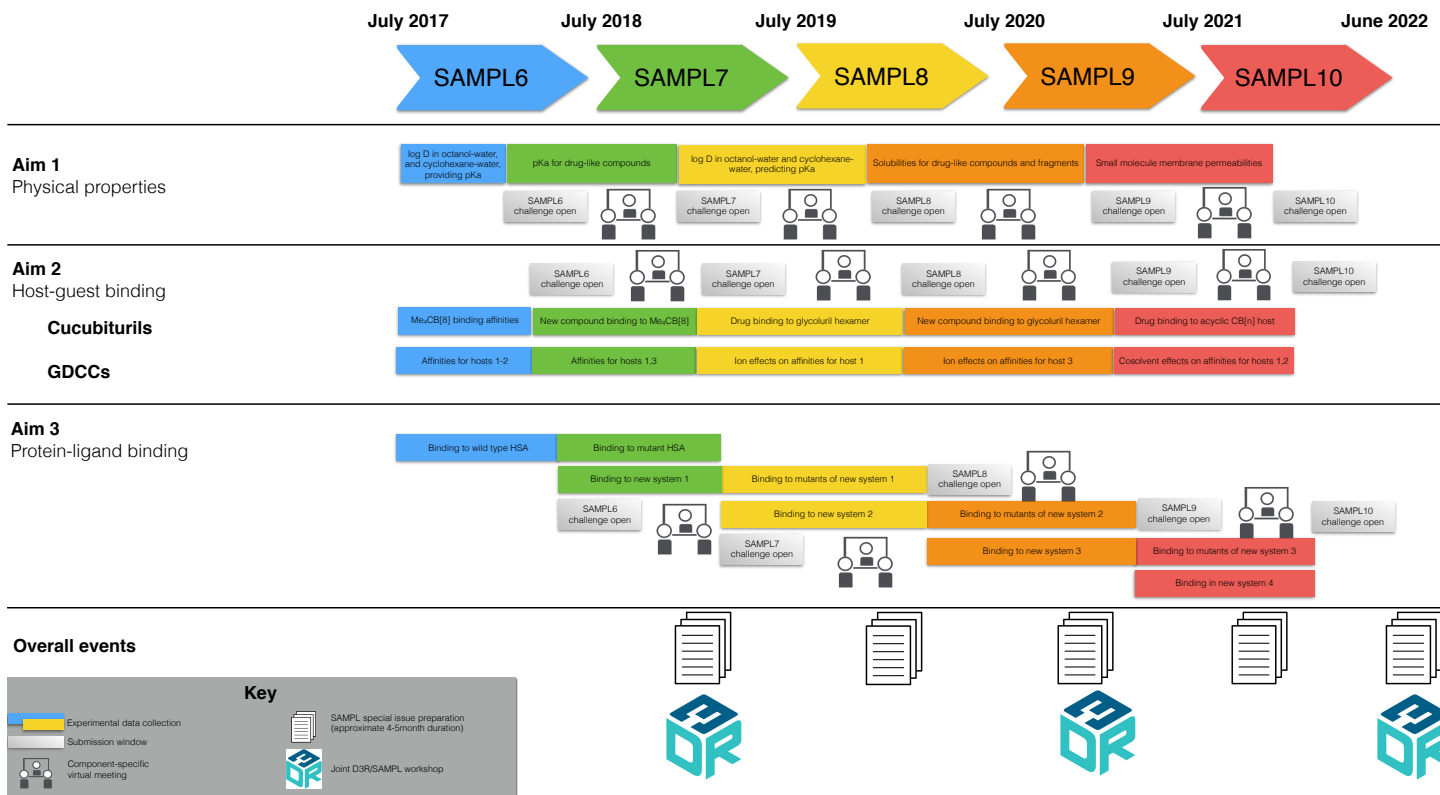
Figure 5. **Timeline for our activities.** Activities covered by this grant include data collection and SAMPL challenges on our three major components (physical properties, host-guest binding, and protein-ligand binding), with each challenge cycle color-coded separately. Data collection within each Aim is shown by a colored bar indicating what is measured and curated. Data collection/curation is followed by a submission window for that challenge component, then all results and analysis are returned to participants and posted on the SAMPL website; this also will nucleate more detailed discussion on the relevant Slack channel (which is expected to already be underway as participants prepare during the data collection phase). Each component will then wrap up with a focused virtual meeting where participants focus on lessons learned and areas which need further exploration; these will be recorded and posted on our website to assist in rapid dissemination of new insights. Virtual meetings precede the submission window for the next SAMPL challenge, giving participants the opportunity to incorporate lessons learned for the next challenge. Submission windows and virtual meetings are staggered across categories so that participants can be involved in all three major areas without multiple simultaneous deadlines. In-person meetings are co-hosted with D3R and will occur every two years. Yearly special issues of JCAMD will have submission deadline shortly after the virtual meeting on the protein-ligand challenge for that year, and an approximate 4-5 month timeline (based on historical experience) from the submission deadline until the special issue appears (with the first papers appearing online substantially sooner). Rapid dissemination of insights is critical for rapid progress, so we highly encourage the use of preprints and informal reports to supplement the special issue.

Each dataset will have a life cycle of collection, curation, blind challenges, and public dissemination. In the past, the unfunded nature of SAMPL has forced us to primarily emphasize the *blind challenges* and pre-challenge *curation* aspects, with isolated forays into collection [39, 92]. This work considers the full life cycle for the first time, with Aims 1–3 dealing primarily with collection and pre-challenge curation. Post-challenge, datasets will receive additional curation and then be released to the community as standard test or benchmark sets that allow retrospective evaluation of methodologies on high-quality data [43]. The FreeSolv dataset, for example, includes a large number of calculated and experimental hydration free energies from SAMPL0–4, and provides a standard benchmark dataset for hydration free energy calculations containing the majority of available high-quality data [**?**]. Post-challenge curation will receive new attention here; in the past, lack of resources has always prevented follow-up experimental work, even when the data clearly indicated it was warranted (such as puzzling issues with dynamic range for logD values in SAMPL5 [31, 39]). The requested budget will ensure that these experiments will now be conducted, allowing computation to drive collection of additional experimental data—potentially by different methodologies—when warranted. Dissemination is the final stage in the data life cycle (see Resource Sharing Plan for details); our driving philosophy is to make the data (including primary data, processed data, and our analysis of challenge submissions) available freely and publicly with permanent, citeable DOIs; ensure relevant data is also deposited in standard community repositories (e.g. BindingDB [**?**]); and guarantee data longevity via backup hosting on library archival facilities (such as UCI's DASH [link]).

Because prior SAMPL work focused mainly on challenges themselves, dissemination of the data has never been a major focus. As part of this project, we will retrieve old SAMPL experimental data, submissions, and analysis from where it is archived and make it available along with the new data generated here.

We will also push for containerization of tools and methods in conjunction with other efforts such as AutoDesk's Molecular Design Toolkit and the NSF Molecular Sciences (MolSci) initiative. Our vision is that long-term, instead of participants submitting a set of predictions, they would also submit the entire workflow they applied via Docker containers in a way which would allow others to reproduce their work if needed, or to repurpose the workflow for new applications, ensuring not just that SAMPL *results* get disseminated, but that the *methods* and *workflows* themselves spread.

COLLABORATION MANAGEMENT PLAN

We have a strong previous history of successful collaboration, with Mobley and Chodera having co-authored roughly a dozen publications since 2006, as well as several workshops and other initiatives. Mobley, Isaacs, and Gibb have also worked together to coordinate past SAMPL challenges, and Mobley and Gibb a previous NSF workshop. PI Mobley will oversee the entire project, with Mobley and Chodera working together on Aim 1 (Mobley working with west coast pharma; Chodera with east coast), Isaacs responsible for Aim 2.1, Gibb for Aim 2.2, and Chodera for Aim 3. Mobley and Chodera will conduct aim 4, involving the other co-investigators as needed. Meetings will consist of a Google Hangout monthly and an in-purpose planning meeting once yearly which will take place at the SAMPL workshops during the years those are offered, and a separate meeting doing the alternate two years. Chodera and Mobley will communicate more frequently due to the interlinked nature of their work. Publications are expected to be largely dictated by the overall Timeline noted above, with an experimental publication associated with each challenge component being prepared for distribution to participants along with their results. Conflict resolution is expected to be straightforward given the extensive past history of interaction and collaboration between the investigators, but if any serious difficulties arise, Michael Gilson (UCSD) will be consulted as an arbiter, given the need this initiative has for close connections with D3R.

OUTLOOK

Physical methods have been slow to achieve their promise in binding prediction, in part because truly significant innovations are so hard to recognize due to a lack of standard tests and benchmarks, and in part because of an "applications first" approach which seems to plague our community where we rush to apply our methods to problems of pharmaceutical relevance without ensuring they can tackle simpler, better-understood problems first. Here, we propose an innovative extension of the successful series of SAMPL blind challenges, generating novel experimental data to drive improvement of the methods in our field and help them become pharmaceutically relevant – beginning with relatively simple physical property prediction and progressing to challenging problems in biomolecular recognition via a series of carefully designed intermediate steps. SAMPL already has a strong track record of success, and funding will ensure an even greater impact on the community and that it is around as a valuable resource for years to come. The proposed series of carefully tailored challenges will focus our community on a variety of problems which we *can* realistically resolve in the near term, resulting in dramatic improvements in the field's ability to do computational molecular design.

Bibliography and References Cited

[1] Mobley, D. L. and Klimovich, P. V.: Perspective: Alchemical free energy calculations for drug discovery. J. Chem. Phys. 137(23): 230901, January 2012.

[2] Christ, C. D. and Fox, T.: Accuracy Assessment and Automation of Free Energy Calculations for Drug Design. J. Chem. Inf. Model. 54(1): 108–120, January 2014.

[3] Deng, N., Forli, S., He, P., Perryman, A., Wickstrom, L., Vijayan, R. S. K., Tiefenbrunn, T., Stout, D., Gallicchio, E., Olson, A. J., and Levy, R. M.: Distinguishing Binders from False Positives by Free Energy Calculations: Fragment Screening Against the Flap Site of HIV Protease. J. Phys. Chem. B. 119(3): 976–988, January 2015.

[4] Sherborne, Bradley, : Opening the lid on FEP. J Comput Aided Mol Des. 2016.

[5] Wang, L., Wu, Y., Deng, Y., Kim, B., Pierce, L., Krilov, G., Lupyan, D., Robinson, S., Dahlgren, M. K., Greenwood, J., Romero, D. L., Masse, C., Knight, J. L., Steinbrecher, T., Beuming, T., Damm, W., Harder, E., Sherman, W., Brewer, M., Wester, R., Murcko, M., Frye, L., Farid, R., Lin, T., Mobley, D. L., Jorgensen, W. L., Berne, B. J., Friesner, R. A., and Abel, R.: Accurate and Reliable Prediction of Relative Ligand Binding Potency in Prospective Drug Discovery by Way of a Modern Free-Energy Calculation Protocol and Force Field. J Am Chem Soc. 137(7): 2695–2703, February 2015.

[6] Christ, C. D. Binding affinity prediction from molecular simulations: A new standard method in structure-based drug design?, May 2016.

[7] Cui, G. Affinity Predictions with FEP+: A Different Perspective on Performance and Utility, May 2016.

[8] Verras, A. Free Energy Perturbation at Merck: Benchmarking against Faster Methods, May 2016.

[9] Schnieders, M. J., Baltrusaitis, J., Shi, Y., Chattree, G., Zheng, L., Yang, W., and Ren, P.: The Structure, Thermodynamics, and Solubility of Organic Crystals from Simulation with a Polarizable Force Field. J. Chem. Theory Comput. 8(5): 1721–1736, May 2012.

[10] Park, J., Nessler, I., McClain, B., Macikenas, D., Baltrusaitis, J., and Schnieders, M. J.: Absolute Organic Crystal Thermodynamics: Growth of the Asymmetric Unit into a Crystal via Alchemy. J. Chem. Theory Comput. 10(7): 2781–2791, July 2014.

[11] Liu, S., Cao, S., Hoang, K., Young, K. L., Paluch, A. S., and Mobley, D. L.: Using MD Simulations To Calculate How Solvents Modulate Solubility. Journal of Chemical Theory and Computation. 12(4): 1930–1941, February 2016.

[12] Leonis, G., Steinbrecher, T., and Papadopoulos, M. G.: A Contribution to the Drug Resistance Mechanism of Darunavir, Amprenavir, Indinavir, and Saquinavir Complexes with HIV-1 Protease Due to Flap Mutation I50V: A Systematic MM–PBSA and Thermodynamic Integration Study. J. Chem. Inf. Model. 53(8): 2141–2153, August 2013.

[13] Lee, C. T., Comer, J., Herndon, C., Leung, N., Pavlova, A., Swift, R. V., Tung, C., Rowley, C. N., Amaro, R. E., Chipot, C., Wang, Y., and Gumbart, J. C.: Simulation-Based Approaches for Determining Membrane Permeability of Small Compounds. J. Chem. Inf. Model. 56(4): 721–733, April 2016.

[14] Comer, J., Schulten, K., and Chipot, C.: Calculation of Lipid-Bilayer Permeabilities Using an Average Force. J Chem. Theory Comput. 10(2): 554–564, February 2014.

[15] Liu, S., Wu, Y., Lin, T., Abel, R., Redmann, J. P., Summa, C. M., Jaber, V. R., Lim, N. M., and Mobley, D. L.: Lead optimization mapper: Automating free energy calculations for lead optimization. J Comput Aided Mol Des. 27(9): 755–770, September 2013.

[16] Mikulskis, P., Genheden, S., and Ryde, U.: A Large-Scale Test of Free-Energy Simulation Estimates of Protein–Ligand Binding Affinities. J. Chem. Inf. Model. 54(10): 2794–2806, October 2014.

[17] Homeyer, N., Stoll, F., Hillisch, A., and Gohlke, H.: Binding Free Energy Calculations for Lead Optimization: Assessment of Their Accuracy in an Industrial Drug Design Context. J. Chem. Theory Comput. 10(8): 3331–3344, August 2014.

[18] Shirts, M. R., Mobley, D. L., and Brown, S. P. Free-energy calculations in structure-based drug design. In Merz, , Kenneth MJ., Ringe, D., and Reynolds, C. H. (Eds.): Drug Design: Structure and Ligand-Based Approaches. Cambridge University Press, January 2010.

[19] Lim, N. M., Wang, L., Abel, R., and Mobley, D. L.: Sensitivity in binding free energies due to protein reorganization. Journal of Chemical Theory and Computation. July 2016.

[20] Rocklin, G. J., Boyce, S. E., Fischer, M., Fish, I., Mobley, D. L., Shoichet, B. K., and Dill, K. A.: Blind Prediction of Charged Ligand Binding Affinities in a Model Binding Site. J. Mol. Biol. 425(22): 4569–4583, November 2013.

[21] Mobley, D. L., Graves, A. P., Chodera, J. D., McReynolds, A. C., Shoichet, B. K., and Dill, K. A.: Predicting absolute ligand binding free energies to a simple model site. J. Mol. Biol. 371(4): 1118–1134, August 2007.

[22] Boyce, S. E., Mobley, D. L., Rocklin, G. J., Graves, A. P., Dill, K. A., and Shoichet, B. K.: Predicting ligand binding affinity with alchemical free energy methods in a polar model binding site. J. Mol. Biol. 394(4): 747–763, December 2009.

[23] Gathiaka, S., Liu, S., Chiu, M., Yang, H., Stuckey, J. A., Kang, Y. N., Delproposto, J., Dunbar, J. B., Carlson, H. A., Burley, S., Walters, W., Amaro, R. E., Feher, V., and Gilson, M. K.: D3R Grand Challenge 2015: Evaluation of Protein-Ligand Pose and Affinity Prediction. J Comput Aided Mol Des. 2016.

[24] Nicholls, A., Mobley, D. L., Guthrie, J. P., Chodera, J. D., Bayly, C. I., Cooper, M. D., and Pande, V. S.: Predicting Small-Molecule Solvation Free Energies: An Informal Blind Test for Computational Chemistry. J. Med. Chem. 51(4): 769–779, February 2008.

[25] Nicholls, A., Wlodek, S., and Grant, J. A.: The SAMP1 Solvation Challenge: Further Lessons Regarding the Pitfalls of Parametrization. J. Phys. Chem. B. 113(14): 4521–4532, April 2009.

[26] Mobley, D. L., Bayly, C. I., Cooper, M. D., and Dill, K. A.: Predictions of Hydration Free Energies from All-Atom Molecular Dynamics Simulations. J Phys Chem B. 113: 4533–4537, January 2009.

[27] Geballe, M. T., Skillman, A. G., Nicholls, A., Guthrie, J. P., and Taylor, P. J.: The SAMPL2 blind prediction challenge: Introduction and overview. J Comput Aided Mol Des. 24(4): 259–279, May 2010.

[28] Geballe, M. T. and Guthrie, J. P.: The SAMPL3 blind prediction challenge: Transfer energy overview. J Comput Aided Mol Des. 26(5): 489–496, April 2012.

[29] Mobley, D. L., Wymer, K. L., Lim, N. M., and Guthrie, J. P.: Blind prediction of solvation free energies from the SAMPL4 challenge. J Comput Aided Mol Des. 28(3): 135–150, March 2014.

[30] Muddana, H. S., Fenley, A. T., Mobley, D. L., and Gilson, M. K.: The SAMPL4 host–guest blind prediction challenge: An overview. J Comput Aided Mol Des. 28(4): 305–317, March 2014.

[31] Bannan, C. C., Burley, K. H., Chiu, M., Shirts, M. R., Gilson, M. K., and Mobley, D. L.: Blind prediction of cyclohexane–water distribution coefficients from the SAMPL5 challenge. J Comput Aided Mol Des. pp 1–18, September 2016.

[32] Yin, J., Henriksen, N. M., Slochower, D. R., Chiu, M. W., Mobley, D. L., and Gilson, M. K.: Overview of the SAMPL5 Host-Guest Challenge: Are We Doing Better? J Comput Aided Mol Des. 2016.

[33] Ignjatović, M. M., Caldararu, O., Dong, G., Muñoz-Gutierrez, C., Adasme-Carreño, F., and Ryde, U.: Binding-affinity predictions of HSP90 in the D3R Grand Challenge 2015 with docking, MM/GBSA, QM/MM, and free-energy simulations. J Comput Aided Mol Des. pp 1–24, August 2016.

[34] Deng, N., Flynn, W. F., Xia, J., Vijayan, R. S. K., Zhang, B., He, P., Mentes, A., Gallicchio, E., and Levy, R. M.: Large scale free energy calculations for blind predictions of protein–ligand binding: The D3R Grand Challenge 2015. J Comput Aided Mol Des. pp 1–9, August 2016.

[35] Sunseri, J., Ragoza, M., Collins, J., and Koes, D. R.: A D3R prospective evaluation of machine learning for protein-ligand scoring. J Comput Aided Mol Des. pp 1–11, September 2016.

[36] Hanson, S. M., Ekins, S., and Chodera, J. D.: Modeling error in experimental assays using the bootstrap principle: Understanding discrepancies between assays using different dispensing technologies. Journal of Computer-Aided Molecular Design. 29(12): 1073–1086, December 2015.

[37] Paranahewage, S. S., Gierhart, C. S., and Fennell, C. J.: Predicting water-to-cyclohexane partitioning of the SAMPL5 molecules using dielectric balancing of force fields. J Comput Aided Mol Des. pp 1–7, August 2016.

[38] Klamt, A., Eckert, F., Reinisch, J., and Wichmann, K.: Prediction of cyclohexane-water distribution coefficients with COSMO-RS on the SAMPL5 data set. J Comput Aided Mol Des. pp 1–9, July 2016.

[39] Rustenburg, A. S., Dancer, J., Lin, B., Feng, J. A., Ortwine, D. F., Mobley, D. L., and Chodera, J. D.: Measuring experimental cyclohexane-water distribution coefficients for the SAMPL5 challenge. bioRxiv. 063081 pp, July 2016.

[40] Bannan, C. C., Calabró, G., Kyu, D. Y., and Mobley, D. L.: Calculating Partition Coefficients of Small Molecules in Octanol/Water and Cyclohexane/Water. Journal of Chemical Theory and Computation. 12(8): 4015–4024, August 2016.

[41] Kollman, P. A.: Advances and continuing challenges in achieving realistic and predictive simulations of the properties of organic and biological molecules. Accounts of Chemical Research. 29(10): 461–469, 1996.

[42] Bhatnagar, N., Kamath, G., and Potoff, J. J.: Prediction of 1-octanol–water and air–water partition coefficients for nitro-aromatic compounds from molecular dynamics simulations. Physical Chemistry Chemical Physics. 15(17): 6467, 2013.

[43] Mobley, D. L. and Gilson, M. K.: Predicting binding free energies: Frontiers and benchmarks. bioRxiv. 074625 pp, September 2016.

[44] Bhakat, S. and Söderhjelm, P.: Resolving the problem of trapped water in binding cavities: Prediction of host-guest binding free energies in the SAMPL5 challenge by funnel metadynamics. J Comput Aided Mol

Des. 2016.

[45] Henriksen, N. M., Fenley, A. T., and Gilson, M. K.: Computational Calorimetry: High-Precision Calculation of Host–Guest Binding Thermodynamics. Journal of Chemical Theory and Computation. 11(9): 4377–4394, September 2015.

[46] Yin, J., Fenley, A. T., Henriksen, N. M., and Gilson, M. K.: Toward Improved Force-Field Accuracy through Sensitivity Analysis of Host-Guest Binding Thermodynamics. The Journal of Physical Chemistry B. 119(32): 10145–10155, August 2015.

[47] Ma, D., Glassenberg, R., Ghosh, S., Zavalij, P. Y., and Isaacs, L.: Acyclic cucurbituril congener binds to local anaesthetics. Supramolecular Chemistry. 24(5): 325–332, May 2012.

[48] Cao, L. and Isaacs, L.: Absolute and relative binding affinity of cucurbit[7]uril towards a series of cationic guests. Supramolecular Chemistry. 26(3-4): 251–258, March 2014.

[49] She, N., Moncelet, D., Gilberg, L., Lu, X., Sindelar, V., Briken, V., and Isaacs, L.: Glycoluril-Derived Molecular Clips are Potent and Selective Receptors for Cationic Dyes in Water. Chem. Eur. J. pp n/a–n/a, August 2016.

[50] Cao, L., Šekutor, M., Zavalij, P. Y., Mlinarić-Majerski, K., Glaser, R., and Isaacs, L.: Cucurbit[7]uril·Guest Pair with an Attomolar Dissociation Constant. Angew. Chem. Int. Ed. 53(4): 988–993, January 2014.

[51] Liu, S., Ruspic, C., Mukhopadhyay, P., Chakrabarti, S., Zavalij, P. Y., and Isaacs, L.: The Cucurbit[n]uril Family: Prime Components for Self-Sorting Systems. Journal of the American Chemical Society. 127(45): 15959–15967, November 2005.

[52] Mock, W. L. and Shih, N. Y.: Structure and selectivity in host-guest complexes of cucurbituril. The Journal of Organic Chemistry. 51(23): 4440–4446, November 1986.

[53] Assaf, K. I. and Nau, W. M.: Cucurbiturils: From synthesis to high-affinity binding and catalysis. Chem Soc Rev. 44(2): 394–418, January 2015.

[54] Moghaddam, S., Yang, C., Rekharsky, M., Ko, Y. H., Kim, K., Inoue, Y., and Gilson, M. K.: New Ultrahigh Affinity Host-Guest Complexes of Cucurbit[7]uril with Bicyclo[2.2.2]octane and Adamantane Guests: Thermodynamic Analysis and Evaluation of M2 Affinity Calculations. Journal of the American Chemical Society. 133(10): 3570–3581, March 2011.

[55] Shetty, D., Khedkar, J. K., Park, K. M., and Kim, K.: Can we beat the biotin–avidin pair?: cucurbit[7]uril-based ultrahigh affinity host–guest complexes and their applications. Chem. Soc. Rev. 44(23): 8747–8761, 2015.

[56] Biedermann, F., Uzunova, V. D., Scherman, O. A., Nau, W. M., and De Simone, A.: Release of High-Energy Water as an Essential Driving Force for the High-Affinity Binding of Cucurbit[n]urils. J. Am. Chem. Soc. 134(37): 15318–15323, September 2012.

[57] Isaacs, L.: Stimuli Responsive Systems Constructed Using Cucurbit[n]uril-Type Molecular Containers. Acc. Chem. Res. 47(7): 2052–2062, July 2014.

[58] Vinciguerra, B., Zavalij, P. Y., and Isaacs, L.: Synthesis and Recognition Properties of Cucurbit[8]uril Derivatives. Org. Lett. 17(20): 5068–5071, October 2015.

[59] Lucas, D., Minami, T., Iannuzzi, G., Cao, L., Wittenberg, J. B., Anzenbacher, P., and Isaacs, L.: Templated Synthesis of Glycoluril Hexamer and Monofunctionalized Cucurbit[6]uril Derivatives. J. Am. Chem. Soc. 133(44): 17966–17976, November 2011.

[60] Ma, D., Zhang, B., Hoffmann, U., Sundrup, M. G., Eikermann, M., and Isaacs, L.: Acyclic Cucurbit[n]uril-Type Molecular Containers Bind Neuromuscular Blocking Agents In Vitro and Reverse Neuromuscular Block In Vivo. Angew. Chem. Int. Ed. 51(45): 11358–11362, November 2012.

[61] Ma, D., Hettiarachchi, G., Nguyen, D., Zhang, B., Wittenberg, J. B., Zavalij, P. Y., Briken, V., and Isaacs, L.: Acyclic cucurbit[n]uril molecular containers enhance the solubility and bioactivity of poorly soluble pharmaceuticals. Nat Chem. 4(6): 503–510, June 2012.

[62] Zhang, B. and Isaacs, L.: Acyclic Cucurbit[n]uril-type Molecular Containers: Influence of Aromatic Walls on their Function as Solubilizing Excipients for Insoluble Drugs. J. Med. Chem. 57(22): 9554–9563, November 2014.

[63] Gilberg, L., Zhang, B., Zavalij, P. Y., Sindelar, V., and Isaacs, L.: Acyclic cucurbit[n]uril-type molecular containers: Influence of glycoluril oligomer length on their function as solubilizing agents. Org. Biomol. Chem. 13(13): 4041–4050, 2015.

[64] Sigwalt, D., Moncelet, D., Falcinelli, S., Mandadapu, V., Zavalij, P. Y., Day, A., Briken, V., and Isaacs, L.: Acyclic Cucurbit[n]uril-Type Molecular Containers: Influence of Linker Length on Their Function as Solubilizing Agents. ChemMedChem. 11(9): 980–989, May 2016.

[65] Zhang, B., Zavalij, P. Y., and Isaacs, L.: Acyclic CB[n]-type molecular containers: Effect of solubilizing group on their function as solubilizing excipients. Org. Biomol. Chem. 12(15): 2413–2422, 2014.

[66] Ma, D., Zavalij, P. Y., and Isaacs, L.: Acyclic Cucurbit[n]uril Congeners Are High Affinity Hosts. J. Org. Chem.

75(14): 4786–4795, July 2010.

[67] Ko, Y. H., Kim, E., Hwang, I., and Kim, K.: Supramolecular assemblies built with host-stabilized charge-transfer interactions. Chem. Commun. (13): 1305–1315, 2007.

[68] Barrow, S. J., Kasera, S., Rowland, M. J., del Barrio, J., and Scherman, O. A.: Cucurbituril-Based Molecular Recognition. Chem. Rev. 115(22): 12320–12406, November 2015.

[69] Urbach, A. R. and Ramalingam, V.: Molecular Recognition of Amino Acids, Peptides, and Proteins by Cucurbit[n]uril Receptors. Isr. J. Chem. 51(5-6): 664–678, May 2011.

[70] Connors, K. A.: Binding Constants. New York, NY, John Wiley & Sons, 1987.

[71] Masson, E., Ling, X., Joseph, R., Kyeremeh-Mensah, L., and Lu, X.: Cucurbituril chemistry: A tale of supramolecular success. RSC Adv. 2(4): 1213–1247, 2012.

[72] Márquez, C., Hudgins, R. R., and Nau, W. M.: Mechanism of Host-Guest Complexation by Cucurbituril. J. Am. Chem. Soc. 126(18): 5806–5816, May 2004.

[73] Biedermann, F., Nau, W. M., and Schneider, H.-J.: The Hydrophobic Effect Revisited—Studies with Supramolecular Complexes Imply High-Energy Water as a Noncovalent Driving Force. Angew. Chem. Int. Ed. 53(42): 11158–11171, October 2014.

[74] 'il Saleh, N., Koner, A., and Nau, W.: Activation and Stabilization of Drugs by Supramolecular pKa Shifts: Drug-Delivery Applications Tailored for Cucurbiturils. Angewandte Chemie. 120(29): 5478–5481, July 2008.

[75] Nau, W. M., Florea, M., and Assaf, K. I.: Deep Inside Cucurbiturils: Physical Properties and Volumes of their Inner Cavity Determine the Hydrophobic Driving Force for Host–Guest Complexation. Isr. J. Chem. 51(5-6): 559–577, May 2011.

[76] Ghosh, I. and Nau, W. M.: The strategic use of supramolecular pKa shifts to enhance the bioavailability of drugs. Advanced Drug Delivery Reviews. 64(9): 764–783, June 2012.

[77] Muddana, H. S., Varnado, C. D., Bielawski, C. W., Urbach, A. R., Isaacs, L., Geballe, M. T., and Gilson, M. K.: Blind prediction of host–guest binding affinities: A new SAMPL3 challenge. J Comput Aided Mol Des. 26(5): 475–487, February 2012.

[78] Gibb, C. L. D. and Gibb, B. C.: Binding of cyclic carboxylates to octa-acid deep-cavity cavitand. J Comput Aided Mol Des. 28(4): 319–325, November 2013.

[79] Sullivan, M. R., Sokkalingam, P., Nguyen, T., Donahue, J. P., and Gibb, B. C.: Binding of carboxylate and trimethylammonium salts to octa-acid and TEMOA deep-cavity cavitands. J Comput Aided Mol Des. pp 1–8, July 2016.

[80] Yin, J., Henriksen, N. M., Slochower, D. R., Shirts, M. R., Chiu, M. W., Mobley, D. L., and Gilson, M. K.: Overview of the SAMPL5 host–guest challenge: Are we doing better? J Comput Aided Mol Des. pp 1–19, September 2016.

[81] Carnegie, R. S., Gibb, C. L. D., and Gibb, B. C.: Anion Complexation and The Hofmeister Effect. Angew. Chem. 126(43): 11682–11684, October 2014.

[82] Lin, Y.-L., Meng, Y., Jiang, W., and Roux, B.: Explaining why Gleevec is a specific and potent inhibitor of Abl kinase. Proc. Natl. Acad. Sci. 110(5): 1664–1669, January 2013.

[83] Shan, Y., Seeliger, M. A., Eastwood, M. P., Frank, F., Xu, H., Jensen, M. O., Dror, R. O., Kuriyan, J., and Shaw, D. E.: A conserved protonation-dependent switch controls drug binding in the Abl kinase. PNAS. 106(1): 139–144, June 2009.

[84] Szakács, Z., Béni, S., Varga, Z., Örfi, L., Kéri, G., and Noszál, B.: Acid-Base Profiling of Imatinib (Gleevec) and Its Fragments. Journal of Medicinal Chemistry. 48(1): 249–255, January 2005.

[85] Grante, I., Actins, A., and Orola, L.: Protonation effects on the UV/Vis absorption spectra of imatinib: A theoretical and experimental study. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy. 129: 326–332, August 2014.

[86] Fasano, M., Curry, S., Terreno, E., Galliano, M., Fanali, G., Narciso, P., Notari, S., and Ascenzi, P.: The extraordinary ligand binding properties of human serum albumin. IUBMB Life (International Union of Biochemistry and Molecular Biology: Life). 57(12): 787–796, December 2005.

[87] Hall, M. L., Jorgensen, W. L., and Whitehead, L.: Automated Ligand- and Structure-Based Protocol for in Silico Prediction of Human Serum Albumin Binding. Journal of Chemical Information and Modeling. 53(4): 907–922, April 2013.

[88] Ràfols, C., Zarza, S., and Bosch, E.: Molecular interactions between some non-steroidal anti-inflammatory drugs (NSAIDŒşs) and bovine (BSA) or human (HSA) serum albumin estimated by means of isothermal titration calorimetry (ITC) and frontal analysis capillary electrophoresis (FA/CE). Talanta. 130: 241–250, December 2014.

[89] Bou-Abdallah, F., Sprague, S. E., Smith, B. M., and Giffune, T. R.: Binding thermodynamics of Diclofenac and Naproxen with human and bovine serum albumins: A calorimetric and spectroscopic study. The Journal

of Chemical Thermodynamics. 103: 299–309, December 2016.

[90] Onufriev, A. V. and Alexov, E.: Protonation and pK changes in protein–ligand binding. Quarterly Reviews of Biophysics. 46(02): 181–209, May 2013.

[91] Merski, M., Fischer, M., Balius, T. E., Eidam, O., and Shoichet, B. K.: Homologous ligands accommodated by discrete conformations of a buried cavity. PNAS. 112(16): 5039–5044, April 2015.

[92] Newman, J., Dolezal, O., Fazio, V., Caradoc-Davies, T., and Peat, T. S.: The DINGO dataset: A comprehensive set of data for the SAMPL challenge. J Comput Aided Mol Des. 26(5): 497–503, December 2011.

[93] Rocklin, G. J., Mobley, D. L., Dill, K. A., and Hünenberger, P. H.: Calculating the binding free energies of charged species based on explicit-solvent simulations employing lattice-sum methods: An accurate correction scheme for electrostatic finite-size effects. The Journal of Chemical Physics. 139(18): 184103, 2013.

[94] Reif, M. M. and Oostenbrink, C.: Net charge changes in the calculation of relative ligand-binding free energies via classical atomistic molecular dynamics simulation. Journal of Computational Chemistry. 35(3): 227–243, January 2014.

[95] Elinder, M., Geitmann, M., Gossas, T., Kallblad, P., Winquist, J., Nordstrom, H., Hamalainen, M., and Danielson, U. H.: Experimental Validation of a Fragment Library for Lead Discovery Using SPR Biosensor Technology. Journal of Biomolecular Screening. 16(1): 15–25, January 2011.

[96] Gilson, M., Given, J., Bush, B., and McCammon, J.: The statistical-thermodynamic basis for computation of binding affinities: A critical review. Biophysical Journal. 72(3): 1047–1069, March 1997.

[97] Latta, M., Knapp, M., Sarmientos, P., Bréfort, G., Becquart, J., Guerrier, L., Jung, G., and Mayaux, J.-F.: Synthesis and Purification of Mature Human Serum Albumin from E. Coli. Bio/Technology. 5(12): 1309–1314, December 1987.

[98] Lang, B. E. and Cole, K. D.: Unfolding properties of recombinant human serum albumin products are due to bioprocessing steps. Biotechnology Progress. 31(1): 62–69, January 2015.

[99] Martin, Y. C.: Let's not forget tautomers. Journal of Computer-Aided Molecular Design. 23(10): 693–704, October 2009.

[100] Temperini, C., Cecchi, A., Scozzafava, A., and Supuran, C. T.: Carbonic anhydrase inhibitors. Comparison of chlorthalidone, indapamide, trichloromethiazide, and furosemide X-ray crystal structures in adducts with isozyme II, when several water molecules make the difference. Bioorganic & Medicinal Chemistry. 17(3): 1214–1221, February 2009.

[101] Rocklin, G. J., Boyce, S. E., Fischer, M., Fish, I., Mobley, D. L., Shoichet, B. K., and Dill, K. A.: Blind Prediction of Charged Ligand Binding Affinities in a Model Binding Site. Journal of Molecular Biology. 425(22): 4569–4583, November 2013.

[102] Venkataraman, H., Verkade-Vreeker, M. C., Capoferri, L., Geerke, D. P., Vermeulen, N. P., and Commandeur, J. N.: Application of engineered cytochrome P450 mutants as biocatalysts for the synthesis of benzylic and aromatic metabolites of fenamic acid NSAIDs. Bioorganic & Medicinal Chemistry. 22(20): 5613–5620, October 2014.

[103] Gunasekaran, K. and Nussinov, R.: How Different are Structurally Flexible and Rigid Binding Sites? Sequence and Structural Features Discriminating Proteins that Do and Do not Undergo Conformational Change upon Ligand Binding. Journal of Molecular Biology. 365(1): 257–273, January 2007.

[104] Khunnawutmanotham, N., Chimnoi, N., Thitithanyanont, A., Saparpakorn, P., Choowongkomon, K., Pungpo, P., Hannongbua, S., and Techasakul, S.: Dipyridodiazepinone derivatives; synthesis and anti HIV-1 activity. Beilstein Journal of Organic Chemistry. 5, July 2009.

[105] Aleksandrov, A., Proft, J., Hinrichs, W., and Simonson, T.: Protonation Patterns in Tetracycline:Tet Repressor Recognition: Simulations and Experiments. ChemBioChem. 8(6): 675–685, April 2007.

[106] Song, Y., Mao, J., and Gunner, M. R.: MCCE2: Improving protein pKa calculations with extensive side chain rotamer sampling. Journal of Computational Chemistry. pp NA–NA, 2009.

[107] Reinhard, L., Mayerhofer, H., Geerlof, A., Mueller-Dieckmann, J., and Weiss, M. S.: Optimization of protein buffer cocktails using Thermofluor. Acta Crystallographica Section F Structural Biology and Crystallization Communications. 69(2): 209–214, February 2013.